Possibilistic contextual skylines with incomplete preferences

Allel Hadjali, Olivier Pivert, Henri Prade

Possibilistic Contextual Skylines with Incomplete Preferences

A. Hadjali
ENSSAT/IRISA, University of Rennes1
22305 Lannion Cedex, France
hadjali@enssat.fr

O. Pivert
ENSSAT/IRISA, University of Rennes1
22305 Lannion Cedex, France
pivert@enssat.fr

H. Prade
IRIT/CNRS, University of Toulouse 3
31062 Toulouse Cedex 09, France
prade@irit.fr

***Abstract*—We propose a possibility theory-based approach to the treatment of missing user preferences in skyline queries. To compensate this lack of knowledge, we show how a set of plausible preferences suitable for the current context can be derived either in a case-based reasoning manner, or using an extended possibilistic logic setting. Uncertain dominance relationships are defined in a possibilistic way and the notion of possibilistic contextual skyline is introduced. This kind of skyline allows us to return the tuples that are non-dominated with a high certainty. The paper also includes a structured overview of the different types of "fuzzy" skylines.**

***Keywords*-Contextual preferences; skyline; possibility theory;**

I. Introduction

In the last two decades, there has been a growing interest in preferences queries in the database community. Numerous approaches have been proposed to make database systems more flexible in supporting user preferences (see [1] for a survey). One of the most well-known approaches is that of skyline queries proposed in [2]. Skyline queries are a popular and powerful paradigm for extracting interesting objects from a multi-dimensional dataset. Given a set r of n-dimensional tuples or points, a skyline query returns the set of non-dominated points in r. A tuple t dominates a tuple t' if t is at least as good as t' in all dimensions and strictly better than t' in at least one dimension.

Several research efforts have been made to develop efficient algorithms and to introduce different variants for skyline queries [3–7]. In particular, the problem of skyline rigidity is also addressed in [8] where a flexible dominance relationship is proposed. It allows enlarging the skyline with points that are not much dominated by any other point (even if strictly speaking they are dominated).

Moreover, in many cases, the expression of preferences inside queries may strongly depend on the context. The term 'context' refers here to any information that can be used to characterize the current query situation [9]. Recently, contextual preferences have received considerable attention in databases and information systems [10–14].

In such works, users explicitly state their preferences together with the queries according to a context. In this paper, we consider the problem of missing information in users' preferences, i.e., how to deal with skyline queries when preferences are not specified for some specific context. To handle such a missing information situation, a set of plausible preferences based on the ones known in other contexts should be derived. In the following, possibility theory [15] is used to model the uncertainty associated with preferences. The proposed approach may be paralleled with the probabilistic method studied in [16]. We use here possibility theory due to its ability to deal with incomplete information, and its qualitative nature that makes it suitable for the problem addressed.

The paper is structured as follows. Section 2 provides a general discussion of the different ways in which a skyline may become "fuzzy" when it is refined, relaxed, simplified, extended to uncertain data, or generalized to incompletely stated context-dependent preferences. The rest of the paper focuses on this latter topic. Section 3 restates the problem using an illustrative example, recalls the probabilistic approach before motivating the approach, and discussing the notion of possibilistic contextual skyline. Section 4 details the approach on the example. Finally, Section 5 enlarges the proposed approach by discussing it in an extended possibilistic setting, before concluding.

II. Different Types of Fuzzy Skylines

There may be many different motivations for making skylines fuzzy in a way or another. First, one may want to refine the skyline by introducing some ordering between its points in order to single out the most interesting ones. Second, one may like to make it more flexible by adding points that strictly speaking do not belong to it, but are close to belonging to it. Third, one may try to simplify the skyline either by granulating the scales of the criteria, or by considering that some criteria are less important than others, or even that some criteria compensate each other, which may enable us to cluster points that are somewhat similar. Fourth, the skyline may be fuzzy due to the uncertainty or the imprecision present in the data. Lastly, the preference ordering on some criteria may depend on the context, and may be specified only for some particular or typical contexts. We now briefly review each of these ideas.

The notion of a skyline in a set of tuples is easy to state (since it amounts to exhibit non dominated points in the sense of Pareto ordering). Namely, assume we have:

- a given a set of criteria $c_1, \cdots, c_n (n \geq 2)$ associated respectively with a set of attributes $i = 1, \cdots, n$;
- a complete ordering $\succcurlyeq_i$ given for each criterion i expressing preference between attribute values[1] (the case of non comparable values is left aside).

A tuple $u = (u_1, \cdots, u_n)$ in a database D belongs to the skyline S if there is no other tuple $u' = (u'_1, \cdots, u'_n)$ in D s.t. for all i, $u'_i \succcurlyeq_i u_i$ and it exists j s.t. $u'_j \succ_j u_j$. In other words, $\forall u' \neq u, \exists i$ s.t. $u'_i \prec_i u_i$ or $\forall j, u'_j \preccurlyeq_j u_j$. Note that the case $u = u'$ is excluded. This means in practice that if two tuples (with different names) are equal componentwise (e.g. two hotels satisfying the considered criteria exactly in the same way), they should be clustered in an equivalence class. Thus, u belongs to the skyline S if $u' \neq u$, $\exists i$ s.t. $u'_i \prec_i u_i$ (then any tuple u' is either dominated by u, or u is non comparable w.r.t. u').

The first idea stated above corresponds to refine the skyline S by stating that u belongs to the fuzzy skyline S_{MP} if $\forall u' \neq u$, $\exists i$ s.t. u_i is much preferred to u'_i. This can be translated in fuzzy set terms by

$$S_{MP}(u) = min_{u' \in D, u' \neq u} max_i MP_i(u_i, u'_i)$$

where MP_i is a fuzzy relation expressing to what extent its first argument is much preferred to the second one for values taken in the domain of attribute i. Clearly, $MP_i(u_i, u'_i) \leq \succ_i (u_i, u'_i)$ where $\succ_i (u_i, u'_i) = 1$ if $u_i \succ_i u'_i$, and is equal to 0 otherwise. Thus, we have $S_{MP} \subseteq S$. $S_{MP} = S$ when MP_i reduces to the crisp relation $\succ_i$. Clearly, S_{MP} may be non normalized, or even empty.

A still more refined fuzzy skyline S^* amounts to select those tuples, if any, that are such that $\forall u' \neq u$, $\exists i$ s.t. u_i is much preferred to u'_i, i.e, they are in S_{MP} to a high degree, and for which $\forall j \in [1, n]$ and $j \neq i$, $\nexists u'$ s.t. u'_j is much preferred to u_j. Thus,

$$\begin{aligned}S^*(u) &= min(min_{u' \in U} max_i MP_i(u_i, u'_i), \\ &\quad min_{j, j \neq i} min_{u' \in U} 1 - MP_j(u'_j, u_j)) \\ &= min_{u' \in U}(max_i MP_i(u_i, u'_i), min_{j, j \neq i} 1 - MP_j(u'_j, u_j)),\end{aligned}$$

with $U = \{u' \in D, u' \neq u\}$. Clearly, $S^* \subseteq S_{MP}$. S^* gathers the most interesting points, since they are much better than the others on at least one attribute, and not so bad on the other attributes (w.r.t. other existing points). To the best of our knowledge, S^* and S_{MP} do not seem to have been considered in the literature.

Rather than refining the skyline, a second type of fuzzy skyline corresponds to the idea of relaxing it, i.e., u still belongs to the skyline to some extent (but to a less extent), if u is only weakly dominated by any other u'. Note first that this does not amount to weakening the requirement $\exists i$ s.t. $u'_i \prec_i u_i$, replacing $\prec_i$ by $max(\prec_i, E_i)$ where E_i is an approximate indifference relation, as $\succ_i$ has been strengthened for defining S_{MP}. Indeed, this would allow for cases where $\exists i$ s.t. u'_i and u_i are approximately equally preferred, while u' dominates u on the rest of the attributes. Rather, such a relaxed skyline S_{REL} can be defined as the extent $S_{REL}(u)$ to which it is false that there exists a tuple u' much preferred to u w.r.t. all attributes (this expression was proposed in [8])

$$\begin{aligned}S_{REL}(u) &= 1 - max_{u' \in U} min_i MP_i(u'_i, u_i) \\ &= min_{u' \in U} max_i (1 - MP_i(u'_i, u_i)).\end{aligned}$$

Note the formal similarity with the expression of S_{MP}, changing $MP_i(u_i, u'_i)$ into $1 - MP_i(u'_i, u_i)$, but mind that u'_i and u_i are interchanged. Clearly, $S \subseteq S_{REL}$, since $1 - MP_i(u'_i, u_i) \geq \preccurlyeq_i (u'_i, u_i)$. If $u \in S$, $S_{REL}(u) = 1$.

Another way of relaxing S is to consider that u still belongs to a fuzzily extended skyline S_{FE} if u is close to u' with $u' \in S$. This would lead to the following definition

$$S_{FE}(u) = max_{u' \in D, u' \neq u} min(S(u'), min_i E_i(u_i, u'_i))$$

where $S(u') = 1$ if $u' \in S$ and 0 otherwise. $S_{FE}(u) = 1$ if $u \in S$ also. Still, one may have $S_{FE}(u) > 0$ while $S_{REL}(u) = 0$. Indeed, a point may be close to a skyline point while being much dominated by another skyline point (e.g. in case of two criteria, a pair (x, y) may be close to a skyline point (x', y) if x is less preferred than x' but close to it, while there exists (x, y') where y' is much preferred to y). Such a situation corresponds to a lack of local "smoothness" of the skyline. So, S_{FE} may be found too permissive.

On the contrary, it may be desirable to simplify the skyline just because it contains too many points [3, 4]. There are many ways to do it. The definitions of S^* and S_{MP} serve this purpose, but they may be empty as already said. We now briefly mention three different ways to simplify the skyline, which are also meaningful. First, one may consider that the set of criteria is partitioned in subsets of decreasing importance. Then we may judge that a tuple cannot belong to the skyline only because it strictly dominates all the other tuples on a non fully important criterion. Indeed it may look strange that a tuple belongs to the skyline while it is dominated on all the important criteria, even if its value on a secondary criterion makes the tuple finally incomparable. In this view, less important criteria may be only used to get rid of tuples that are dominated on immediately less important criteria, in case of ties on more important criteria. Note that however the importance of a criterion might be changed in case of a very strong (and very rare) dominance for a non fully important criterion.

A second, completely different idea for simplifying a skyline is to use coarser scales for the evaluation of the attributes (e.g., moving from precise values to rounded values). This will lead to more comparable (or even identical) tuples. Still another way to increase the number of comparable tuples is to use a *2-discrimin* (or more generally an order *k-discrimin*) ordering [17]. This can be illustrated

[1] $u \succ v$ means u is preferred to v. $u \succcurlyeq v$ means u is at least as good as v, i.e., $u \succcurlyeq v \Leftrightarrow u \succ v \vee u \approx v$, where $\approx$ denotes indifference.

by the following example, where we compare hotels on the basis of their price, distance to the station, and distance to a conference location. Then $(80, 1, 3)$ et $(70, 3, 1)$ are not Pareto comparable, while we may consider that the two distance criteria play similar roles and that there is equivalence between the sub-tuples $(1, 3)$ and $(3, 1)$ leading to compare the tuples on the remaining components. Order *k-discrimin* will allow for equivalences between sub-tuples of size k or less.

The fourth type of "fuzzy" skyline is quite clear. When attributes values are imprecisely or more generally fuzzily known, we are led to define the tuples that certainly belong to the skyline, and those which only possibly belong to the skyline, using necessity and possibility measures [7]. In the following we concentrate on the last category of "fuzzy" skyline that is induced by an incompletely known context-dependency of the involved preferences.

III. Uncertain Contextual Skylines

A. A Motivating Example

We use an example taken from [16]. It consists of a relation with 3 attributes *Price*, *Distance* and *Amenity* about a set of hotels (Table I). For instance, a skyline query returns "those hotels for which there is *no cheaper* and, at the same time, *closer* to the beach alternative". One can easily check that the skyline contains hotels h_4 and h_5. In other terms, hotels h_4 and h_5 represent non-dominated hotels w.r.t. *Price* and *Distance* dimensions. In many cases, results of preference queries may depend on the context. To this end, we consider contextual skyline queries that require that the dominance relationships will be defined relative to a context. To illustrate this idea, consider three contexts C_1, C_2 and C_3 shown in Table II (where a given context can be composed at most by two context parameters (*Purpose*, *Period*)). For example, when the user is on a business trip in June (context C_1), hotels h_3, h_4 and h_5 are the results of the skyline query for C_1. See Table II for contexts C_2 and C_3 and their corresponding skylines.

Table I: Relation describing hotels

Hotel	Price	Distance	Amenity
h_1	200	10	Pool(P)
h_2	300	10	Spa(S)
h_3	400	15	Internet'I)
h_4	200	5	Gym(G)
h_5	100	20	Internet(I)

Table II: Contextual Skylines

Context	Preferences	Skyline
C_1: Business, June	$I \succ G, I \succ \{P, S\}$ $G \succ P, S$	h_3, h_4, h_5
C_2: Vacation	$S \succ \{P, I, G\}$	h_2, h_4, h_5
C_3: Summer	$P \succ \{I, G\}$ $S \succ \{I, G\}$	h_1, h_2, h_4, h_5
C_q: Business, Summer	-	?

Let us now examine situation C_q (fourth row in Table II), where the user plans a business trip in the summer but states no preferences. To understand the uncertainty present in preferences in the context C_q, let us consider amenities *Internet* (I) and *Pool* (P). One can observe that: (i) I may be preferred to P as in C_1; (ii) P may be preferred to I as in C_3, or (iii) I and P may be equally favorable as in C_2. In fact, all three alternatives hold with some plausibility that depends on the similarity of C_q to C_1, C_2, and C_3. Moreover, the uncertainty propagates to the dominance relationships, i.e., every hotel may dominate another with a certainty degree that depends on the context.

B. Contextual skylines

Let D be a relational database with attributes $A_1, \cdots, A_n$. Let C be a context associated with the user query. C is modeled as a set of parameter-value pairs. E.g., context C_1 of Table II is expressed as $Purpose = Business, Period = June$. Let $dom(A_j)$ be the domain of attribute A_j and $u, v \in dom(A_j)$. We use the notation $u \succ_{A_j} v$ to express that u is preferred to v for attribute A_j. This implies that for two database tuples t and t', if $t.A_j = u, t'.A_j = v, u \succ_{A_j} v$, and $t.A_k \approx t'.A_k$ for all $k \neq j$, then t is preferred to t'. Now, if preferences depend on context, we denote by $u \succ_{A_j} v|C_i$ a preference on attribute A_j that holds for a particular context C_i.

Example (Cont'). Let us come back to our previous example and consider the *Amenity* attribute (denoted by A). The following contextual preferences hold (see Table II): $I \succ_A G|C_1$, $I \succ_A P|C_1$, $I \succ_A S|C_1$, $G \succ_A P|C_1$ and $G \succ_A S|C_1$. One can observe that the set of preferences for attribute A_j in context C_i defines a (strict) partial order. The *Contextual Skyline Query* (CSQ) returns the database tuples that are not dominated by any other in context C_i. Let t and t' two tuples, t dominates t' in C_i, noted $t \succ_A t'|C_i$,

$$\text{iff } \forall j, t.A_j \succcurlyeq t'.A_j|C_i \text{ and } \exists k, t.A_k \succ t'.A_k|C_i.$$

C. Problem Description

Let the *profile* of a user refer to the set of contextual preferences specified in the past. Now, given the current context C_q for which preferences are missing, and given the user's profile, the problem is how to determine the *non-dominated tuples* in context C_q.

In the case where C_q perfectly matches only one of the contexts included in the profile, the problem reduces to a precise CSQ. The difficulty arises when C_q is not present in the profile. In this case, what should the user's preferences be? A solution is to interpolate them on the basis of the profile. This entails that derived preferences will be pervaded with uncertainty, i.e., for two tuples t and t', the relation t *dominates* t' holds only with some *plausibility degree*.

In [16] an approach is proposed where uncertainty is modeled using probability theory. The probability that t

dominates t' in context C_q is defined as[2] $Pr(t \succ t' \mid C_q) = \prod_j Pr(t.A_j \succcurlyeq_{A_j} t'.A_j \mid C_q)$ if $t \neq t'$ and 0 otherwise, with $Pr(u \succ_{A_j} v \mid C_q) = \sum_k Sim(C_q, C_k) \cdot |u \succ_{A_j} v|C_k| / \sum_k Sim(C_q, C_k)$ and $u, v \in dom(A_j)$. $Sim(C_q, C_k)$ is the similarity between the contexts C_q and C_k and $|u \succ_{A_j} v|C_k|$ is 1 if such a preference exists, and 0 otherwise. The probability that a tuple t belongs to the skyline is also defined, i.e., $P_{sky}^{C_i} = \prod_{t' \neq t}(1 - Pr(t' \succ t|C_i))$. The major drawback of this approach stems from the fact that the probabilistic assessment of uncertainty should be quantitative, and is hence more demanding in terms of data. In the following, we propose an alternative approach that relies on possibility theory where uncertainty modeling can be qualitative, thus avoiding the need of quantifying uncertainty if information is poor.

D. Possibilistic Contextual Skylines

Here, the possibilistic framework (see Appendix) is used to model uncertainty. To achieve this, each uncertain contextual preference will be associated with a certainty (necessity) degree. We use the following notation:

$$N(u \succcurlyeq_{A_j} v|C_i) = \alpha \tag{1}$$

with $\alpha \in [0, 1]$. This expression reads: in context C_i, u is preferred or equally preferred to v with a certainty α. Since the degree α plays an ordinal role only, the interval $[0, 1]$ can be replaced by any linearly ordered scale. Using the dual relation between necessity measure and possibility measure Π, the following relation holds

$$\Pi(\neg(u \succcurlyeq_{A_j} v|C_i)) = 1 - \alpha \tag{2}$$

The approach proposed relies on a well-known Case-Based Reasoning (CBR) principle [18] which states that "the more similar the values of two situations for attributes s and s_0, the more possible the similarity of their values on their outcome attributes t and t_0". Thus, for each pair $(u, v) \in dom(A_j) \times dom(A_j)$, one can assess the possibility that the relation $u \succ_{A_j} v|C_q$ holds in the following way:

$$\Pi(u \succ_{A_j} v|C_q) = max_i(min(Sim(C_q, C_i), N(u \succ_{A_j} v|C_i))) \tag{3}$$

where $Sim(C_q, C_i)$ is detailed further and $N(u \succ_{A_j} v|C_i)$ is the certainty that $u \succ_{A_j} v$ holds in the context C_i (i.e., $N(u \succ_{A_j} v|C_i) = 1$ if $u \succ_{A_j} v$ is true, and 0 otherwise).

Note that formula (3), which parallels the probabilistic expression, is in fact an empirical adaptation of the above principle. When used in CBR, the principle amounts to the expression of a possibility as a maximum of possibility degrees weighted by similarities. Note that: (i) When $Sim(C_q, C_i) > 0$, $\Pi(u \succ_{A_j} v|C_q) = 0$ if there is no certainty at all that $u \succ_{A_j} v|C_i$ holds (i.e. $N(u \succ_{A_j} v|C_i) = 0$), and $\Pi(u \succ_{A_j} v|C_q) = 1$ if it is certain that $u \succ_{A_j} v|C_i$ holds (i.e., $N(u \succ_{A_j} v|C_i) = 1$) when C_q matches exactly one C_i; (ii) Now, if it is certain that $u \succ_{A_j} v|C_i$ holds in all contexts, one has only $\Pi(u \succ_{A_j} v|C_q) = max_i Sim(C_q, C_i)$.

[2] In [16], $u \succcurlyeq v$ is defined as $u \succ v \vee u = v$.

In a similar way as in (3), one can compute the possibility that the relation $(u \succ_{A_j} v|C_q)$ does not hold, i.e.,

$$\Pi(\neg(u \succ_{A_j} v|C_q)) = max_i(min(Sim(C_q, C_i), N(\neg(u \succ_{A_j} v|C_i)))). \tag{3'}$$

Uncertain dominance relationships. Since preferences are pervaded with uncertainty, the dominance relationship becomes uncertain. Then, for two tuples t and t', the certainty degree that t dominates t' in the context C_i is now defined as:

$$N(t \succ t'|C_i) = \begin{cases} 0 \text{ if } \forall j, t.A_j \approx t'.A_j|C_i \\ min_j N(t.A_j \succcurlyeq_{A_j} t'.A_j|C_i) \text{ otherwise} \end{cases} \tag{4}$$

where the second case applies only when tuples do not have equally preferred values in all attributes of interest.

Classically, a tuple is in the skyline if it is not dominated by any other tuple of the same relation. To adapt this definition to our case, one can compute the skyline certainty of a tuple t in the following way:

$$N_{sky}^{C_i}(t) = min_{t' \neq t} N(\neg(t' \succ t)|C_i) = min_{t' \neq t}\{1 - \Pi(t' \succ t|C_i)\} \tag{5}$$

Formula (5) expresses the certainty that a tuple t is not dominated by any other tuple t' of the same relation. One can easily check that in the deterministic case, when all preferences are totally certain, the scalar $N_{sky}^{C_i}(t)$ is equal to

- 1 if t is not dominated, i.e., $\forall t', N(\neg(t' \succ t)|C_i) = 1$,
- 0 otherwise, i.e., $\exists t', N(\neg(t' \succ t)|C_i) = 0$.

α-Certain Contextual Skyline Query. Given a database D and a set of uncertain preferences, our final aim is to retrieve the tuples t whose skyline certainty is greater or equal to a user-defined threshold α, i.e., compute the set $\alpha\text{–}CCSQ(D/C_i)$ defined as:

$$\alpha\text{–}CCSQ(D/C_i) = \{t \in D, N_{sky}^{C_i}(t) \geq \alpha\} \tag{6}$$

Context Similarity. Let us now say a few words about the similarity relation, Sim, between contexts. Let us consider that each context C contains n context parameters $(P_1, P_2, \ldots, P_n)$ and assume also that for each parameter P_i there exists a function Sim_{P_i} that measures the similarity between two values of its domain $dom(P_i)$. One may use the following functions depending on the domain type: (i) for numerical domains, $Sim_{P_i}(x, y) = 1 - (|x - y|/(M - m))$ where M (resp. m) is the maximum (resp. minimum) value of $dom(P_i)$; (ii) for categorical domains, $Sim_{P_i}(x, y) = |lvs(x) \cap lvs(y)|/|lvs(x) \cup lvs(y)|$, i.e., the Jaccard index where $lvs(x)$ stands for the set of leaves under x; (iii) for nominal domains, $Sim_{P_i}(x, y) = 1$ if $x = y$, 0 otherwise. The similarity measure between two contexts C and C' can be defined as: $Sim(C, C') = min_{k=1,n} Sim_{P_k}(C.P_k, C'.P_k)$.

IV. A Detailed Example

We use the example provided in Section III to illustrate the notions described in the previous section. Let us first estimate the extent to which the query context C_q is similar to the contexts present in the user's profile. Assume that the following similarity relations between parameter values are available: $Sim_{Purpose}(Business, Vacation) = 0$, $Sim_{Period}(June, Summer) = 1/3$. Then, $Sim(C_q, C_1) = 1/3$, $Sim(C_q, C_2) = 0$ and $Sim(C_q, C_3) = 1$.

Now based on formula (3), uncertain preferences in the context C_q can be inferred and are summarized in Table III. Let us provide some details about the calculus of one of the

Table III: Preferences Possibilities $\Pi(u \succ_A v|C_q)$

$u\backslash v$	I	G	P	S
I	-	1/3	1/3	1/3
G	0	-	1/3	1/3
P	1	1	-	0
S	1	1	0	-

degrees of possibility given in Table III: $\Pi(I \succ_A G|C_q) = max(min(Sim(C_q, C_1),\ N(I \succ_A G|C_1)), min(Sim(C_q, C_2), N(I \succ_A G|C_2)), min(Sim(C_q, C_3), N(I \succ_A G|C_3))) = max(min(1/3, 1), min(0, 0), min(1, 0)) = 1/3$.

To assess the skyline certainty of each tuple t of Table I, we need to compute the uncertain dominance relationships among tuples shown in this table. See Table IV for the skyline certainties obtained for all tuples. Let us take an example to illustrate the calculus of dominance certainties of Table IV.

Table IV: Dominance certainties $N(s \succ t|C_q)$

$s\backslash t$	h_1	h_2	h_3	h_4	h_5
h_1	0	0	2/3	0	0
h_2	0	0	2/3	0	0
h_3	0	0	0	0	0
h_4	0	0	0	0	0
h_5	0	0	0	0	0

$N(h_1 \succ h_3|C_q) = min_j N(h_1.Aj \succcurlyeq h_3.Aj|C_q) = min(N(200 \succcurlyeq 400|C_q), N(10 \succcurlyeq 15|C_q), N(P \succcurlyeq I|C_q))$. One can easily see that $N(200 \succcurlyeq 400|C_q)$= $N(10 \succcurlyeq 15|C_q) = 1$. By $N(A \cup B) = max(N(A), N(B))$, we have $N(P \succcurlyeq I|C_q) = max(N(P \succ I|C_q), N(P \approx I|C_q)) = N(P \succ I|C_q)$ since $N(P \approx I|C_q) = 0$. Using formula (3'), $\Pi(\neg(P \succ I|C_q)) = max(min(1/3, 1), min(0, 1), min(1, 0)) = 1/3$. This means that $N(P \succ I|C_q) = 2/3$ and $N(P \succcurlyeq I|C_q) = 2/3$. Thus, $N(h_1 \succ h_3|C_q) = min[1, 1, 2/3] = 2/3$.

We have now to compute the skyline certainty for each tuple t, i.e., $N_{sky}^{C_q}(t)$. To achieve this by applying formula (5), we need the degree of certainty that t is not dominated by another tuple s (i.e., $N(\neg(s \succ t|Cq))$). Table V summarizes those certainty degrees. One can observe that h_1 is not dominated by h_2 for sure (i.e., $N(\neg(h_2 \succ h_1|Cq)) = 1$), while h_1 is not dominated by h_4 with a certainty equal to $2/3$. Let us give details about the calculus of $N(\neg(h_2 \succ h_1|Cq))$. It is easy to see that $N(\neg(h_2 \succ h_1|C_q)) = 1 - \Pi(h_2 \succ h_1|C_q)$. Unfortunately from Table IV, no information can be obtained on $\Pi(h_2 \succ h_1|C_q)$ (we have only $\Pi(\neg(h_2 \succ h_1|C_q)) = 1$ due to $N(h_2 \succ h_1|Cq) = 0$). It is why, we use the following formula: $\Pi(h_2 \succ h_1|C_q) = \Pi(\wedge_j(h_2.A_j \succcurlyeq h_1.A_j)|C_q) = min_j \Pi(h_2.A_j \succcurlyeq h_1.A_j|C_q) = min(0, 1, 0) = 0$ (see Appendix). So, $N(\neg(h_2 \succ h_1|C_q)) = 1$. In a similar way, we compute $N(\neg(h_4 \succ h_1|C_q)) = 1 - \Pi(h_4 \succ h_1|C_q) = 1 - min(1, 1, 1/3) = 2/3$. From Table IV and using the relation $min(N(A), N(A^c)) = 0$, we have $N(\neg(h_1 \succ h_3|C_q)) = 0$.

Table V: Non-dominance certainties $N(\neg(s \succ t|C_q))$

$s\backslash t$	h_1	h_2	h_3	h_4	h_5
h_1	1	1	0	1	1
h_2	1	1	0	1	1
h_3	1	1	1	1	1
h_4	2/3	2/3	1	1	1
h_5	1	1	1	1	1
$N_{sky}^{C_q}(t)$	2/3	2/3	0	1	1

According to formula (5), one can compute the skyline certainty of each hotel h_i of Table I in context C_q, see the last row of Table V. For instance, we have $N_{sky}^{C_q}(h_1) = 2/3$ and $N_{sky}^{C_q}(h_4) = 1$. So, $\alpha\text{–}CCSQ(D/C_q) = \{h_1, h_2, h_4, h_5\}$ for $\alpha = 2/3$. Let us take a look at hotel h_3 which belongs to the skyline with a certainty 0 despite the fact that context C_q is somewhat similar to context C_1 (where h_3 is in the skyline). This can be argued by the fact that C_q is more similar to C_3 than to C_1, and our approach leads to a result which is more similar to the one provided in C_3 (where h_3 is not in the skyline) than in C_1. However, one could check that h_3 belongs to the skyline with a possibility degree equal to $1/3$. Note that the degree $1/3$ have to be interpreted ordinally and not quantitatively.

V. Towards a Possibilistic Logic Handling

In the previous sections, we have proposed a similarity-based handling of context-dependent preferences. Still, such an approach is not fully satisfactory in different situations. First, the context C_q may be in fact a subcontext of one (or several) contexts where the preference orderings is/are known. Then, determining the preferences applicable to C_q is a simple matter of logical inference, not of CBR-like reasoning. Moreover, it may happen that among the contexts where the preferences are known, we have two contexts C and C' such that $C' \models C$ (i.e., C' entails C), with conflicting preferences. This has to be interpreted in a non monotonic logic manner, considering that we use the C-preferences if $C_q \models C$ and $C_q \not\models C'$, and the C'-preferences if $C_q \models C'$. This calls for a logical approach that is now outlined.

Generally, context dependency can be handled in the setting of an extended possibilistic logic, as the one used in [19] for preference queries with uncertain data. Suppose we have one criterion like amenity in the example, with values including a and b (e.g., internet and pool). Using possibilistic logic, the preference of a over b in context C_1 writes $\{(\neg C_1 \vee \neg a(x) \vee Ans(x), 1, \mathbf{1}), (\neg C_1 \vee \neg b(x) \vee Ans(x), \alpha, \mathbf{1})\}$, where $a(x)$ expresses that a is true for item x and $Ans(x)$ that item x is in the set of eligible answers, and α is a priority level, and $\mathbf{1}$ expresses the (full) certainty that e.g., having b in C_1 has a priority level $\alpha < 1$, while the preference of b over a in context C_2 writes $\{(\neg C_2 \vee \neg b(x) \vee Ans(x), 1, \mathbf{1}), (\neg C_2 \vee \neg a(x) \vee Ans(x), \alpha, \mathbf{1})\}$.

Suppose both C' and C_2 are true, where C' is a subcontext of C_1, which writes$(\neg C' \vee C_1, 1, \mathbf{1})$, while 1 is the maximal priority. Then we can deduce $(\neg a(x) \vee Ans(x), 1, \mathbf{1})$ and $(\neg b(x) \vee Ans(x), 1, \mathbf{1})$ by applying the possibilistic resolution rule both on the priority levels and on the certainty levels. Thus opposite preferences are understood disjunctively: here having a is as good as having b.

Assume further, e.g., that we know $a(h_1)$, i.e. hotel h_1 has amenity a. Then, if we just know we are in context C', we can conclude $(Ans(h_1), 1, \mathbf{1})$, which means that h_1 is non-dominated on the Amenity attribute (due to its maximal priority), with full certainty ($\mathbf{1}$). Lastly, if we want to also handle similar contexts, we may here express the counterpart of the CBR principle by stating that if the formula $(\neg C \vee \neg g(x) \vee Ans(x), \lambda, \boldsymbol{\rho})$ holds, then the formula $(\neg C' \vee \neg g(x) \vee Ans(x), \lambda, \boldsymbol{min(\rho, Sim(C, C'))})$ also holds. This expresses that we downgrade the certainty of a preference by the amount of similarity between the contexts.

VI. Conclusion

The paper has provided a structured discussion of the different types of "fuzzy" skylines that may be imagined, before focusing on the case of incompletely specified user preferences in specific contexts. We have shown how to derive a set of plausible preferences suitable for the context at hand on the basis of the information known for other contexts. We use a possibilistic setting for defining uncertain dominance relationships. Then the user is provided with the tuples that are not dominated with a high certainty, leading to a notion of possibilistic contextual skyline.

Appendix. Brief Overview on Possiblity Theory

Possibility theory [15] offers a qualitative model for uncertainty where any event A is characterized by two measures: its possibility Π (expressing the fact that A may more or less occur) and its necessity N (expressing that A will occur more or less for sure). The necessity N of A is defined as: $N(A) = 1 - \Pi(A^c)$ where A^c is the event opposite to A. The following properties are of interest: (i) $max(\Pi(A), \Pi(A^c)) = 1$; (ii) $\Pi(A \cup B) = max(\Pi(A), \Pi(B))$; (iii) $\Pi(A) < 1 \Rightarrow N(A) = 0$; (iv) $N(A \cap B) = min(N(A), N(B))$; (v) $\Pi(A \cap B) = min(\Pi(A), \Pi(B))$ and $N(A \cup B) = max(N(A), N(B))$, if A and B are logically independent.

Possibilistic logic is based on the inference rule:
$N(p \vee q) \geq \alpha$ and $N(\neg p \vee r) \geq \beta \Rightarrow N(q \vee r) \geq min(\alpha, \beta)$.

References

[1] A. Hadjali, S. Kaci, and H. Prade, "Database preferences queries: a possibilistic logic approach with symbolic priorities," in *Proc. FoIKS*, 2008, pp. 291–310.

[2] S. Borzsony, D. Kossmann, and K. Stocker, "The skyline operator," in *Proc. ICDE*, 2001, pp. 421–430.

[3] C.-Y. Chan, H. V. Jagadish, K.-L. Tan, A. K. H. Tung, and Z. Zhang, "Finding k-dominant skylines in high dimensional space," in *ACM SIGMOD 2006*, pp. 503–14.

[4] X. Lin, Y. Yuan, Q. Zhang, and Y. Zhang, "Selecting stars: The k most representative skyline operator," in *Proc. ICDE*, 2007, pp. 86–95.

[5] M. E. Khalefa, M. F. Mokbel, and J. J. Levandoski, "Skyline query processing for incomplete data," *Proc. ICDE*, pp. 556–565, 2008.

[6] J. Pei, B. Jiang, X. Lin, and Y. Yuan, "Probabilistic skylines on uncertain data," in *VLDB 2007*, pp. 15–26.

[7] E. Hüllermeier, I. Vladimirskiy, B. Prados Suárez, and E. Stauch, "Supporting case-based retrieval by similarity skylines: Basic concepts and extensions," in *Proc. ECCBR'08*, 2008, pp. 240–254.

[8] M. Goncalves and L. Tineo, "Fuzzy dominance skyline queries," in *Proc. DEXA*, 2007, pp. 469–478.

[9] A. K. Dey, "Understanding and using context," *Personal Ubiquitous Comput.*, vol. 5, pp. 4–7, 2001.

[10] T. Strang and C. Linnhoff-Popien, "A context modeling survey," in *Workshop on Advanced Context Modelling, Reas. and Manag.*, 2004.

[11] S. Holland and W. Kießling, "Situated preferences and preference repositories for personalized database applications," *Conceptual Modeling*, pp. 511–523, 2004.

[12] A. van Bunningen, "Context aware querying - challenges for data management in ambient intelligence," Tec. Rep., Enschede, Dec. 2004.

[13] K. Stefanidis, E. Pitoura, and P. Vassiliadis, "Adding context to preferences," in *ICDE*, 2007, pp. 846–855.

[14] A. Hadjali, A. Mokhtari, and O. Pivert, "A Fuzzy-Rule-Based Approach to Contextual Preference Queries," in *Proc. IPMU*, 2010, pp. 532–541.

[15] Dubois and Prade, *Possibility Theory*. Plenum 1988.

[16] D. Sacharidis, A. Arvanitis, and T. Sellis, "Probabilistic contextual skylines," in *ICDE*, 2010, pp. 273–284.

[17] H. Prade, "Refinement of Minimum-Based Oredering in between Discrimin and Leximin," in *Proc. Linz Seminar on Fuzzy Set Theory*, 2001, pp. 39–43.

[18] D. Dubois, H. Prade, F. Esteva, P. Garcia, L. Godo, and R. de Màntaras, "Fuzzy set modelling in case-based reasoning," *Int. J. Int. Sys.*, vol. 13, pp. 345–373, 1998.

[19] P. Bosc, O. Pivert, and H. Prade, "A possibilistic logic view of preference queries to an uncertain data base," in *Proc. Fuzz-IEEE*, 2010, pp. 379–384.